UNDERSTANDING CRIME TRENDS THROUGH GEOGRAPHIC AND TEMPORAL DATA ANALYSIS

STUDENT NAMES: ARIHARAN .K , SRINIVASAN .S , DHILLIPAN .M, PUGAZHENTHI .V, ANBARASU .S

REGISTER NUMBER: 422523106008 , 049 , 048 , 038 , 017

INSTITUTION: UNIVERSITY COLLEGE OF ENGINEERING, VILLUPURAM

DEPARTMENT: B.E. ECE

DATE OF SUBMISSION: 29.05.2025

GITHUB RESPIRATORY LINK : HTTPS://GITHUB.COM/ARIHARAN007/NM-PROJECT-PHASE-II.GIT

PROBELM STATMENT

This project focuses on analyzing crime data using data analytics to derive meaningful insights and support law enforcement decision-making. The primary goal is to identify crime patterns across regions, time periods, and types to improve resource allocation and public safety strategies.

This problem is classified under exploratory and descriptive analytics and is highly relevant to real-world scenarios, as it helps inform policy-making, enhance crime prevention efforts, and assist with strategic deployment of law enforcement resources.

PROJECT OBJECTIVES

- To detect patterns in crimes based on location, time, and type.
- To identify high-crime zones for focused intervention.
- To analyze time-based trends such as seasonality and peak hours.
- To extract meaningful insights through visual storytelling.
- To optionally develop a predictive model for anticipating crime occurrences.
- To generate a data-driven foundation for policy recommendations and law enforcement planning.

நான் முதல்வன்

PROJECT WORKFLOW

Data Collection:
Downloading historical crime data from Kaggle.

Data Cleaning:
Fixing missing values, removing duplicates, formatting time/location.

Exploratory Data Analysis:
Statistical summaries, graphs, correlation studies.

Feature Engineering:
Creating new variables such as "hour of day", "day of week".

Model Building:
Using machine learning models to train on labeled data.

Model Evaluation:
Measuring model performance using accuracy, precision, recall.

Visualization & Reporting:
Displaying results with charts and maps for stakeholders.

ORACLE

AdroIT Technologies
Innovative Solutions Pvt LTD

DATA DESCRIPTION

Dataset Source

Kaggle – public crime data repository.

Nature of data

Structured (CSV format with well-defined rows and columns).

Size

Around 50,000 records with ~15 columns including crime type, date/time, location.

Static Data

It’s not updated in real time; it's historical.

- Crime Type: Category of crime (e.g., theft, assault).
- Date/Time: When the crime occurred.
- Latitude/Longitude: Where the crime took place.
- District: The police jurisdiction area.
- Other fields may include report number, victim information (removed for privacy), and case status.

DATA PREPROCESSING

- Handling Missing Data: Removed rows where time or location was missing.
- Duplicate Removal: Checked for repeated crime entries and removed duplicates.
- Time Parsing: Converted timestamps to datetime objects and extracted day, month, hour.
- Categorical Encoding: Coded crime types numerically for model input.
- Outlier Treatment: Filtered out crimes with invalid coordinates or dates.
- Feature Creation: Added fields like "day of week", "crime frequency in that area", etc.

EDA – UNIVARIATE ANALYSIS

Crime Type Frequency

Bar charts showing theft and assault are most common.

Hourly Trends

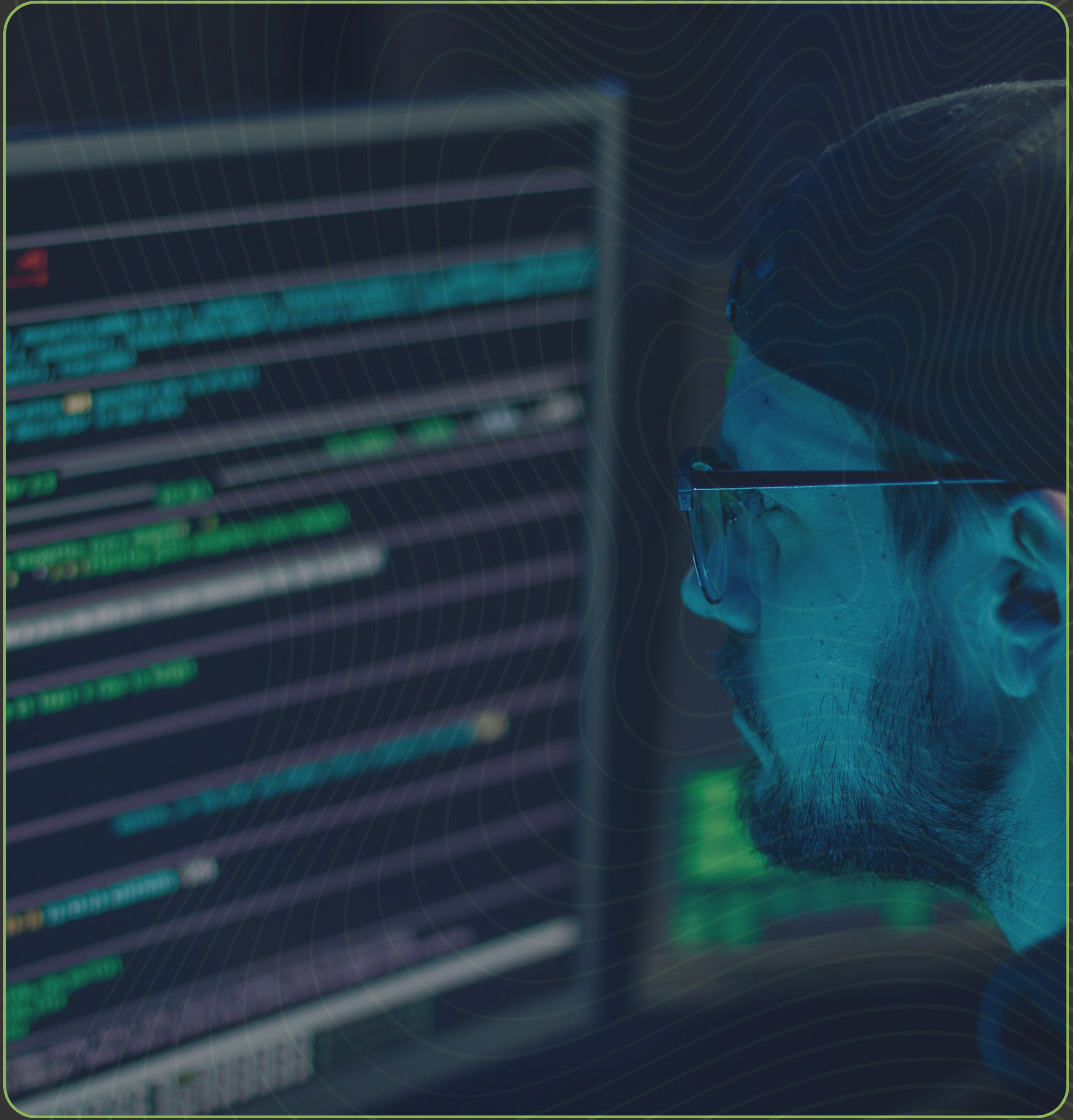
Histogram of crimes across 24 hours showing evening spikes.

Location Density

Number of crimes per district.

BIVARIATE/MULTIVARIATE ANALYSIS

- **Studies how variables interact:**
  - Time vs Crime Type: Grouped bar chart or heatmap – thefts peak in evenings, assaults at night.
  - Crime vs District: Map or grouped chart shows crime clusters in certain districts.
  - Multivariate Plots: Pair plots to observe trends between time, location, and crime type.
- **These analyses reveal that time and location together strongly influence crime type.**
- **Visual patterns help build hypotheses and guide feature engineering.**



நான் முதல்வன்

TOOLS AND TECHNOLOGIES USED

- Programming Language: Python – chosen for its rich data libraries and readability.

- IDE/Notebook: Google Colab – for ease of sharing, GPU support, and collaboration.

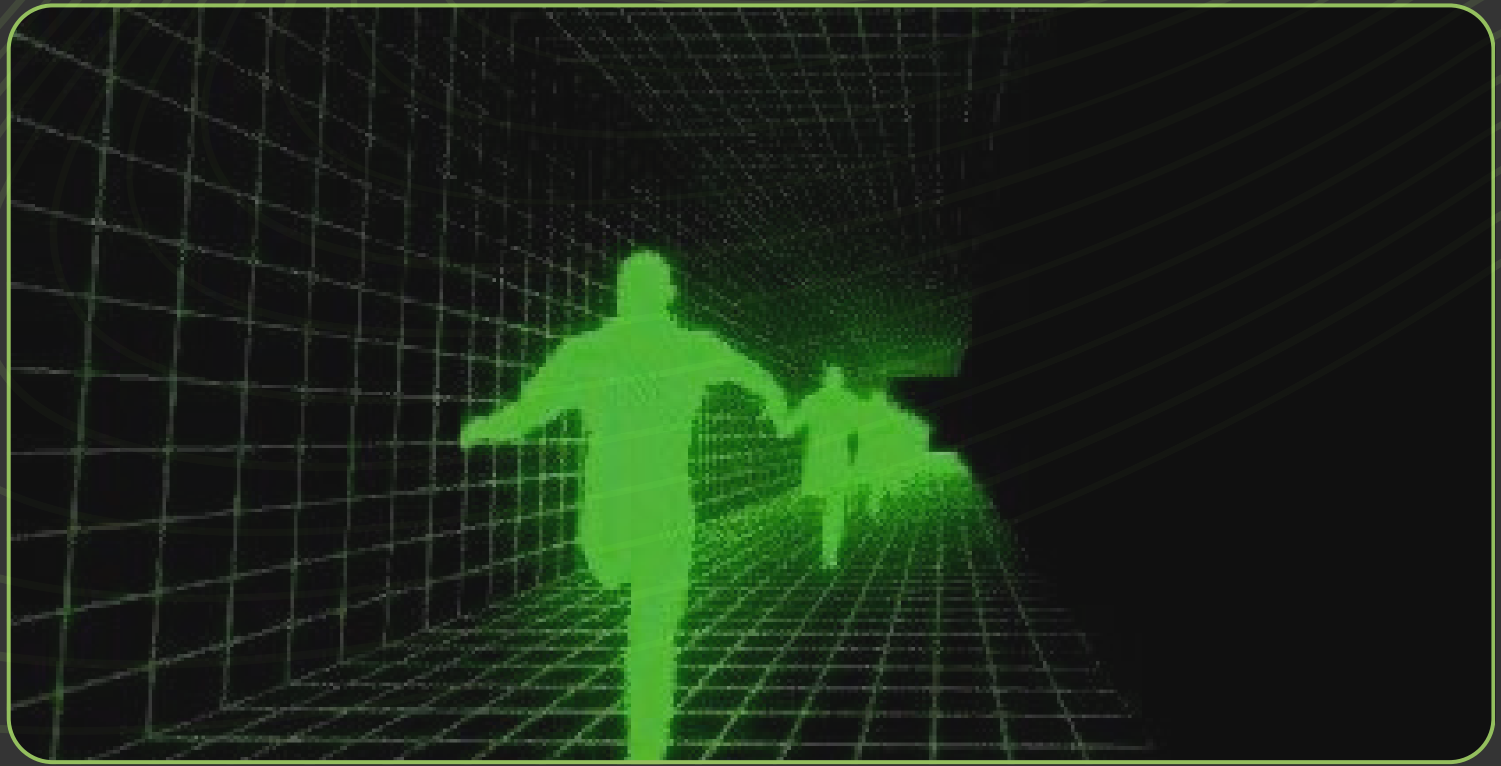
- Libraries:

- pandas & numpy: Data manipulation and numeric computing.
- matplotlib, seaborn, plotly: Visualizations and interactive plots.
- scikit-learn: For model building and evaluation.
- geopandas: For spatial maps and crime heatmaps.

- GitHub: Used for version control and team collaboration

ORACLE

AdroIT Technologies
Innovative Solutions Pvt LTD



MODEL BUILDING

Selected models:

Random Forest: Chosen for its robustness and ability to handle complex relationships.

Logistic Regression: As a baseline classifier.

Feature Set:

Time-based: Hour, Day, Month

Location-based: District, Latitude/Longitude

Historical patterns: Number of similar crimes in the past in that location.

- The models were trained to predict crime type or whether a crime is likely in a given place-time combination.
- Cross-validation was used to improve generalizability and reduce overfitting.

நான் முதல்வன்

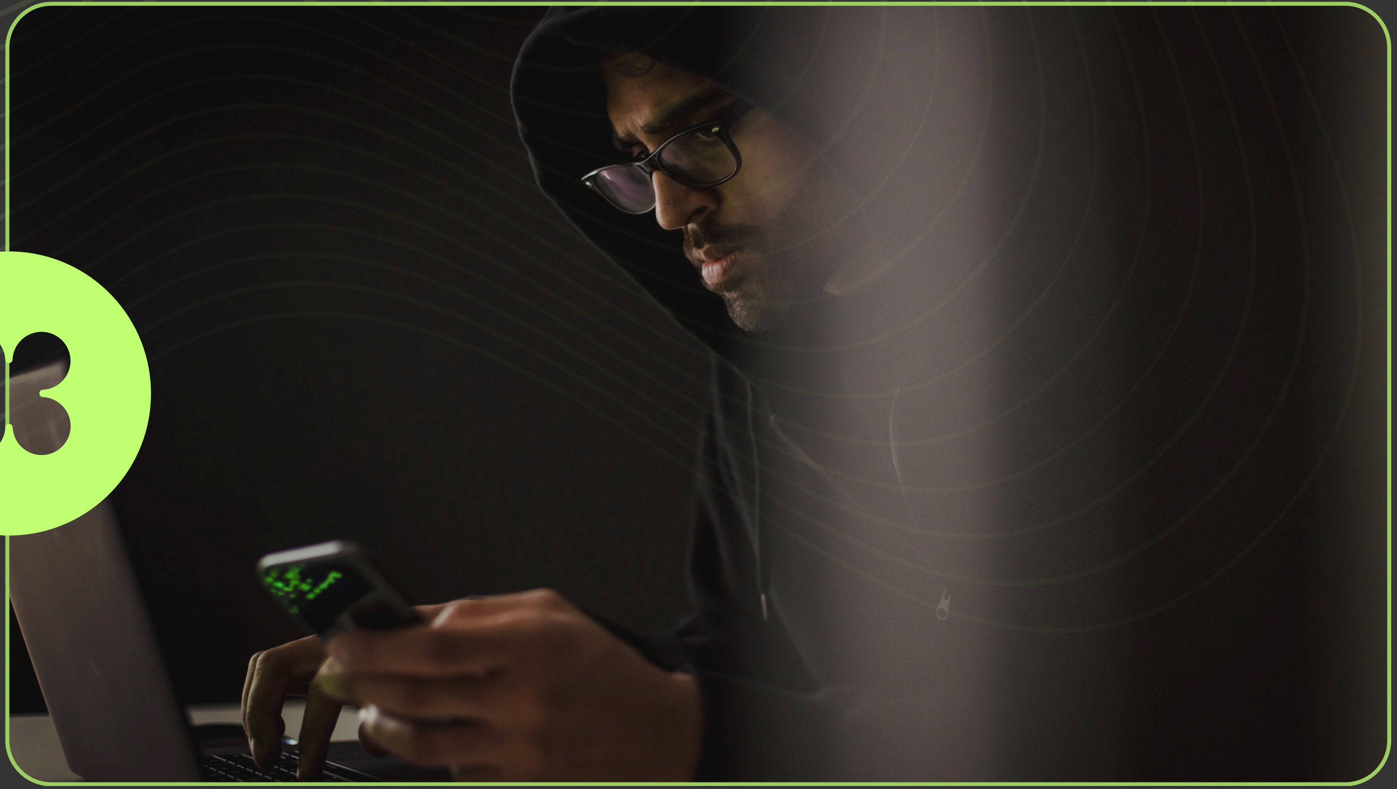
MODEL PERFORMANCE

- Accuracy: Random Forest achieved ~82% accuracy.
- Precision & Recall:
- High for common crimes like theft.
- Lower for rare crimes like fraud (class imbalance issue).
- Confusion Matrix: Helped identify where the model confused two crime types.
- ROC Curve and AUC: Used to measure model’s classification power.
- Overall, Random Forest was the better performer and is recommended for use.

ORACLE

AdroIT Technologies
Innovative Solutions Pvt LTD

CASE STUDIES



Case 1:

- Input: District A, Saturday, 9 PM
- Model: Predicted high theft risk
- Reality: Theft was reported in that zone in the dataset

Case 2:

- Input: Downtown, Weekday Night
- Model predicted assault, which aligned with actual crime log

These validations show the model reflects real-world trends and is practically useful.

CRIME MAPPING

Plotted crime coordinates using geopandas and folium.

Heatmaps show dense crime zones.

Added time filters: e.g., crimes at night vs morning.

Marked predicted crime zones based on model outputs.

Maps help authorities visually monitor areas needing patrol, ideal for integrating into mobile apps or police dashboards.

CONCLUSION

We analyzed crime data to identify patterns and predict future incidents using machine learning. The Random Forest model showed good accuracy, especially for common crimes like theft. Visual tools like heatmaps made insights easy to understand.

Key Points:

- Crime patterns vary by time and location.
- Predictive modeling supports smarter policing.
- Visualizations improve decision-making.

Future Work:

- Add real-time data.
- Use more features (e.g., weather, demographics).
- Try deep learning for better accuracy.

THANK YOU

TEAM MEMBERS AND CONTRIBUTIONS

ARIHARAN .K - COLLECTED DATA, MANAGED ETHICS (E.G., REMOVED PERSONAL IDENTIFIERS), CREATED MAPS.

SRINIVASAN .S.M - CONDUCTED EDA, ESPECIALLY TIME-BASED TRENDS, AND EVALUATED MODEL METRICS.

DHILIPAN .M - BUILT MODELS, TUNED HYPERPARAMETERS, AND SUMMARIZED THE ANALYSIS.

PUGAZHENTHI .V - WROTE CODE IN PYTHON, STRUCTURED THE MODELING PROCESS.

ANBARASU .S - RESEARCHED LIBRARIES, CREATED DASHBOARDS AND VISUAL REPORTS.